Word melodies vs. pitch accents: A perceptual evaluation of terracing contours in British and Nigerian English

Carlos Gussenhoven¹, Inyang Udofot²

¹Department of Linguistics, Radboud University Nijmegen, Netherlands
²Department of English, University of Uyo, Nigeria
c.gussenhoven@qmul.ac.uk, imudofot@gmail.com

Abstract
The results of a perception experiment in which Nigerian English listeners judged the well-formedness of Nigerian English intonation contours suggests that the language has tonal specifications for each syllable, including syllables that are unstressed in British English. The association of pitch accents to accented syllables in British English explains why British English listeners are relatively insensitive to deviations in the pitch of unstressed or unaccented syllables.

Index Terms: perception, intonation, word melody, pitch accent

I. Introduction
The terracing intonation pattern is given as the ‘core [intonation] pattern’ of Nigerian English by [1]. Descending series of pitch peaks were first described for African languages as ‘downdrift’. Later, this concept was merged with what was then termed ‘downstep’, the lowered pronunciation of a H-tone relative to a preceding H-tone without an intervening pitch valley [2; 3]. The term ‘downstep’ is now applied to the ‘staircase’ type as well as to downstepping contours with low valleys between the steps. This is particularly appropriate for Nigerian English, where the descending steps may or may not be interrupted by low pitches due to function words intervening between the lexical words which form the steps.

In our interpretation, the contour is due to a tone grammar intervening between the lexical words which form the steps. We assume that the syllable is the tone bearing unit, as it is in indigenous languages like Ibibio. The last target represents the ‘staircase’ type as well as to downstepping contours with pitch valley [2; 3]. The term ‘downstep’ is now applied to the ‘staircase’ type as well as to downstepping contours with low valleys between the steps. This is particularly appropriate for Nigerian English, where the descending steps may or may not be interrupted by low pitches due to function words intervening between the lexical words which form the steps. In our interpretation, the contour is due to a tone grammar as formulated in [4] according to which each lexical word receives a H-tone and each function word a L-tone. In terms of the typology in [5, ch. 3], in which tones come as (a) word melodies, (b) pitch accents, and (c) boundary tones, the Nigerian tones are monotonal word melodies, which spread through the words they come with. Since they are single tones, no direction or edge needs to be specified. The crucial aspect here is the word domain: the pitch will change at any boundary between a lexical word and a function word (from H to L) and between any two lexical words (from H to a lowered H). Graphically, the downstepping pattern is shown in (1). A second purpose was to provide evidence for the word melody status of tones in the grammar of Nigerian English and of the specification of tone to every syllable, as shown in (1). A second purpose was to provide evidence for the typological difference between Nigerian English and British English, in which only accented syllables are specified for tone through pitch accents and intervening syllables remain unspecified for tone, resulting in a looser connection between f0 and stretches of unaccented syllables.

II. Experiment I: Nigerian English
II-A. Method
In order to test Nigerian listeners’ sensitivity to (a) the L-toned status of function words and (b) the location of the pitch change at word boundaries rather than at specific syllables within words, a corpus of twelve sentences was composed with three naturally accentable words, referred to as W₁, W₂ and W₃. One or two unaccented function words are specified for tone through pitch accents and intervening syllables remain unspecified for tone, resulting in a looser connection between f0 and stretches of unaccented syllables.

![Diagram of pitch accent pattern](image-url)

The function words in (2) can alternatively be pronounced with low pitch in BrE. The variable here is phrasing: an intonational phrase (IP) boundary after *It’s a case will typically attract an initial %L boundary tone before unaccented of.*

The purpose of the experiment was, first, to provide evidence for the word melody status of tones in the grammar of Nigerian English and of the specification of tone to every syllable, as shown in (1). A second purpose was to provide evidence for the typological difference between Nigerian English and British English, in which only accented syllables are specified for tone through pitch accents and intervening syllables remain unspecified for tone, resulting in a looser connection between f0 and stretches of unaccented syllables.

The results of a perception experiment in which Nigerian English listeners judged the well-formedness of Nigerian English intonation contours suggests that the language has tonal specifications for each syllable, including syllables that are unstressed in British English. The association of pitch accents to accented syllables in British English explains why British English listeners are relatively insensitive to deviations in the pitch of unstressed or unaccented syllables.

Index Terms: perception, intonation, word melody, pitch accent

I. Introduction

The terracing intonation pattern is given as the ‘core [intonation] pattern’ of Nigerian English by [1]. Descending series of pitch peaks were first described for African languages as ‘downdrift’. Later, this concept was merged with what was then termed ‘downstep’, the lowered pronunciation of a H-tone relative to a preceding H-tone without an intervening pitch valley [2; 3]. The term ‘downstep’ is now applied to the ‘staircase’ type as well as to downstepping contours with low valleys between the steps. This is particularly appropriate for Nigerian English, where the descending steps may or may not be interrupted by low pitches due to function words intervening between the lexical words which form the steps.

In our interpretation, the contour is due to a tone grammar as formulated in [4] according to which each lexical word receives a H-tone and each function word a L-tone. In terms of the typology in [5, ch. 3], in which tones come as (a) word melodies, (b) pitch accents, and (c) boundary tones, the Nigerian tones are monotonal word melodies, which spread through the words they come with. Since they are single tones, no direction or edge needs to be specified. The crucial aspect here is the word domain: the pitch will change at any boundary between a lexical word and a function word (from H to L) and between any two lexical words (from H to a lowered H). Graphically, the downstepping pattern is shown in (1). A second purpose was to provide evidence for the word melody status of tones in the grammar of Nigerian English and of the specification of tone to every syllable, as shown in (1). A second purpose was to provide evidence for the typological difference between Nigerian English and British English, in which only accented syllables are specified for tone through pitch accents and intervening syllables remain unspecified for tone, resulting in a looser connection between f0 and stretches of unaccented syllables.

II. Experiment I: Nigerian English

II-A. Method

In order to test Nigerian listeners’ sensitivity to (a) the L-toned status of function words and (b) the location of the pitch change at word boundaries rather than at specific syllables within words, a corpus of twelve sentences was composed with three naturally accentable words, referred to as W₁, W₂ and W₃. One or two unaccented function words are specified for tone through pitch accents and intervening syllables remain unspecified for tone, resulting in a looser connection between f0 and stretches of unaccented syllables.
TABLE I
Sentences used in the perception experiment. The third lexical word (W3) has the prosodic structure [wsws] (group A), [wsS] (group B) and [wS] (group C).

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>He’s a member of both associations</td>
<td>Give it more consideration</td>
<td>It’s a case of late assimilation</td>
<td>That’s a strange interpretation</td>
</tr>
<tr>
<td>B</td>
<td>The ball was last September</td>
<td>We start in late October</td>
<td>The men will speak Ibibio</td>
<td>She wants to sing cantatas</td>
</tr>
<tr>
<td>C</td>
<td>They grow the best bananas</td>
<td>We need a new container</td>
<td>He saw his old employer</td>
<td>We’ll wait for the new arrival</td>
</tr>
</tbody>
</table>

Five speakers of standard Nigerian English were asked to read out the corpus in a neutral and quiet voice at least twice. They were recorded using an Olympus VN-2100PC Digital voice recorder in the second author’s office. From these recordings we selected two speakers, one male and one female, whose recordings were technically the best and contained complete sets of sentences. Subsequently, we selected one reading of each sentence from the recordings by each of these two speakers. In general, speakers read the sentences in a careful, deliberate style. In about a fifth of the cases we decided to shorten pauses between words. The resulting file always sounded more natural than the original to the second author, who is a speaker of Nigerian English. One third of the materials, sentences A3, B2 and C1, was additionally recorded by a 30-year-old English female speaker of English in the studio of the School of Languages, Linguistics and Film of Queen Mary, University of London, from which readings one utterances of each sentence was selected.

The selected readings were annotated with the help of Praat [10]. Employing conventional segmentation practice, we identified the CV-boundary of the first syllable of W1, the first function word, W2 and W3, as well as of the main stressed syllable of W3. (Recall that in the case of W1 and W2 the first syllable is also the syllable with word stress.) We also identified all word boundaries. In order to test two specific aspects of the well-formedness of the NigE contour given in 1, we decided to produce four artificial f0 contours on each of the 24 selected utterances produced by the two speakers of NigE and on each of the three selected utterances produced by the speaker of BrEng. Using the Psola resynthesis function in Praat, the f0 of the ‘terrace steps’ for W1, W2 and W3 was 270 Hz, 230 Hz and 200 Hz for the female speakers and 170 Hz, 135 Hz and 110 Hz for the male speaker, for all utterances. In versions (a) and (c), the f0 of W1 continued across the function word(s), while in versions b and d, the f0 dropped to 180 Hz at the CV boundary of the (first) function word, from where it continued to a point 60 ms to the left of beginning of the rhyme of W2. Versions (a) and (b) had a pitch drop after W2, causing the part before the main stress of W3 to be lower in pitch than W2, while versions c and d had this pitch drop at the main stress of W3, causing the pitch of the first part of W3 to be the same as W2. The f0 of the periphery of all sentences, i.e. everything to the left of CV1 minus 60 ms and everything to the right of CV5 plus 60 ms, was left unchanged. All pitch rises and falls were produced in 60 ms. Figure 2 shows the pitch contours used in the resynthesis of the four versions of It’s a case of late assimilation. This procedure led to the generation of 4 × 12 (sentences) × 2 (speakers) or 96 stimuli produced from NigE source utterances, plus 4 × 3 (sentences) or 12 stimuli produced from the BrE source utterances, or 108 stimuli in all.

II-B. Procedure
An audio-cd was prepared containing the 108 stimuli in a random order. Each stimulus was preceded by a 200 ms bleep and an 800 ms pause, and followed by a 3.7 s response period. Four fillers occurred at the beginning of the test and two at the end. The audio-cd was played over loudspeakers in a large room in the Department of English at Uyo University. Judges were given the task to rate each stimulus for the degree to which it conformed to the speech of an educated speaker of Nigerian English when reading a sentence as a radio newsreader or university teacher. They gave their judgements on a printed answer sheet immediately after hearing the stimulus by ticking one of the five boxes numbered 1 to 5, where ‘5’ represented
significant different from each of the other three contours (p<0.01 for I and III and p<0.05 for contour IV), and that additionally contour III is significantly different from contour IV.

II-D. A speaker of British English

In order to see whether the preferences of the Nigerian listeners for the realization of Nigerian English intonation contours are exclusive to speech that is segmentally and rhythmically Nigerian English, a set of 12 stimuli was included which consisted of the four contours shown in Fig 1 superimposed on selected utterances of the sentences He’s a member of both associations, We start in late October and He saw his old employer as spoken by a 30-year old female speaker of English from England. An analysis of variance on the scores by the Nigerian listeners with Contour and Syllable type as factors yielded no significant effects. This finding suggests that the Nigerian intonational preferences as apply to Nigerian English, not to British English. Rather, these stimuli are marked down regardless of contour (a mean score of 2.46, well below the mean score for the Nigerian English speakers, 3.13).

III. Experiment II: Nigerian vs British English

Phonologically, Nigerian English has syllabic tone: every syllable has a specification. There is no reason to assume that any syllables are phonologically privileged and thus phonetically more salient than others. British English, by contrast, has intonational pitch accents, which leave varying stretches of unaccented and/or unstressed syllables unspecified for tone. In this section, we consider the question whether Nigerian listeners show greater sensitivity to pitch on syllables which in British English are unaccented and/or unstressed than British English listeners. The hypothesis is that British English listeners will show little if any differentiation in their acceptability judgements of the four contours. To test this prediction, twenty-three judges were recruited from the student population of Queen Mary, University of London, 9 male and 14 female, 9 in the 18-20 age range and 14 in the 21-25 age range. They were told that they were going
to listen to a number of Nigerian English speakers who had undergone a course in British English intonation, and who were trying their best to produce British English intonation patterns. Judges were asked to indicate on a 5-point scale how well the speaker had succeeded in producing an English intonation pattern in each case.

Out of the $23 \times 108$ judgements, two were missing. The scores were processed in the same way as in Experiment I. An analysis of variance (repeated measures) with Speaker, Syllable Type and Contour as factors yielded significant main effects for all three factors: Speaker ($F(1,22)=11.26$, $p<.01$), Syllable Type ($F(2,22)=5.02$, $p<.05$) and Contour $F(3,22)=2.95$, $p<.05$). Speaker KE was judged to be 0.29 scale point more acceptable than speaker NVO, quite as was the case for the Nigerian judges. Post-hoc Sidak comparisons for Contour showed no significant differences between any pairs, which means that two of the contours were together significantly different from the other two together (see Fig 4).

Syllable type C (bananas for W3) received lower scores that the other two syllable types, which we attribute to the unreduced nature of the word-initial syllable. Pronunciations like /bananas/, /arajval/ are apparently more disturbing to British English listeners than non-reduction in cases like /lasosiejan/. The differentiation among the contours is between those with low pitched function words (contours I and IV) and high-pitched function words (contours I and III). We explain this finding on the basis of the declarative contour in (1). The significant difference between contour III and the less acceptable contour IV shows that failure to lower for the function words is a more salient deviation from the norm than failure to downstep at the word boundary; no significant difference was found between contours I and III, which only differed in the location of the downstep on W3.

### V. Conclusions

Nigerian English intonation is to be described as a word tone system, in which the tones are predictable: H for lexical words, L for function words. Phonetic implementation requires consecutive H’s to be downstepped, regardless of intervening Ls. Quite unlike British English, which places pitch accents on words as a function of a number of grammatical and pragmatic factors, Nigerian English word melodies are inherent components of lexical representations. The observation that Nigerian English does not ‘deaccent’ or has ‘end stress’ is inappropriate to the extent that the notions ‘stress’ and ‘accent’ are inapplicable to the language. Its word tones are not simply negotiable the way pitch accents are in English.

### VI. Acknowledgements

This research was carried out with financial support from the British Academy and the Association of Commonwealth Universities enabling the second author to spend a month in each of 2008 and 2009 at the School of Languages, Linguistics and Film of Queen Mary, University of London (Collaborative Grant DB-2007-03/04 to the first author). We are grateful to Martin Barge, Rachel Fournier, Joop Kerkhoff, Brian Place and Frans van der Slik for various kinds of technical assistance. We thank KE, NVO and Julia Rollitt for recording the sentences used in the experiment.

### References